Top View Human Head and Shoulder Classification Using CNN

Ivan Rivalcoba*†, Isaac Rudomín*, Krely Rodrguez‡
*Barcelona Supercomputing Center, Barcelona, Spain
†Instituto Tecnológico de Gustavo A. Madero, CDMX, México
‡Instituto Tecnológico de Minatitlan, Veracruz, México
E-mail: {ivan.rivalcoba, isaac.rudomin}@bsc.es kyan.itmina@gmail.com

Keywords—Human Detection, Computer Vision, CNN.

I. Extended Abstract

Both industries and scientists consider the human behavior analysis on crowds a powerful source of data for computer applications in the field of security on smart cities and surveillance systems, urban design and planning, video games, autonomous car to mention a few. That is one of the reasons pushing academic and commercial researchers towards a deeper understanding of how humans act, make decisions and plans when they are cruising their environment.

Before even thinking of getting a piece of software capable to analyze a crowd, the first challenge as a priority to tackle is the human detection and tracking. The two main crucial requirements for this topic are high accuracy and real-time speed that means human detectors that are accurate enough to be relied on and fast enough to run on commercial computer hardware, taking as an example of the above mentioned, hardware limited on compute power such as smartphones or tablets. In the present document we focus on the human detection usually the first stage over any approach of human tracking.

Through the years there have been many different types of methods to detect people in video sequences, the first one was the Viola Jones [1] proposed in 2001 who described a machine learning approach for visual object recognition and also introduced the Integral Images", a popular technique to accelerate the calculus of many descriptors. Dalal et al. [2] performed a complete study of Histogram of Oriented Gradients (HOG) applied to the representation of humans. This method offers good results for pedestrian detection by evaluating local histograms of image gradient orientations over a dense normalized overlapping grid, giving a better accuracy in comparison with Viola jones method but with one important drawback and is that it requires a multiscale sliding window causing a bad performance on speed.

All the above methods now are considered as the classical approach to deal with the problem of recognition that was originated by a work published in 2013 by Pierre Sermanet [3], they proposed a multi-scale sliding window algorithm using Convolutional Neural Networks (CNNs). And suddenly after that work, deep learning has become a standard in computer vision tasks [4]. The following section will describe a system developed to classify human head and shoulder as the first stage of a bigger system to model human steering from real humans recorded on video sequences.

Fig. 1. Head and shoulder of humans are Omega shaped.

A. A CNN APPROACH TO CLASSIFY HUMAN HEAD AND SHOULDER REGIONS ON VIDEO SEQUENCES

We present a system capable of classifying head and shoulder sections of a human. We decided to detect the head and shoulders section of the body due to the fact that the human head remains constant over all the scene, unlike other parts of the body. The head and shoulders section are omega (Ω) shaped, see Figure 1. This property makes the head and shoulders the most stable parts of the body to be detected and tracked.

We propose the following CNN architecture designed to prevent the use of huge amount of data to train the neural network (Figure 2).

It is worth to notice that we avoid using aggressive dropout, finding a good balance to avoid overfitting but saving training time. The Dataset used to train the network is the same employed by Li et al. [5], this dataset consist of 3909 images for training and 2143 images for validation, to ensure a better generalization of the omega shape, we added data augmentation to generate more images for our training. The augmentation consisted in the inclusion of transformations including rotation, horizontal and vertical flip. The detailed

TABLE I. Data augmentation allows to get a plausible train with a limited dataset

Transformation	Value
Rotation range	30
Width shift range	0.1
Height Shift Range	0.1
Shear Range	0.2
Zoom Range	0.2
Horizontal flip	TRUE

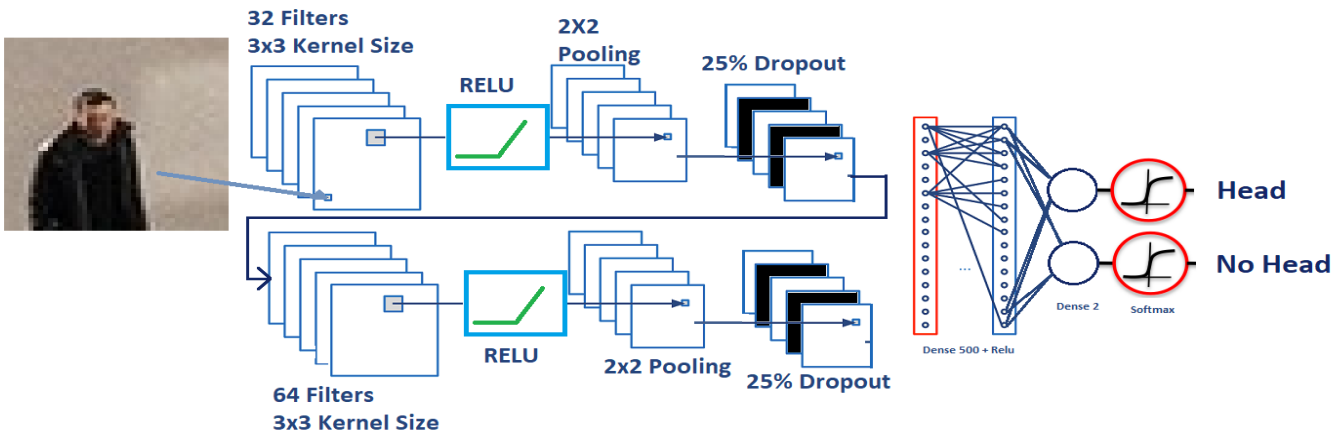


Fig. 2. Architecture of Convolutional Neural Network performing the classification task

transformations are showed in table I.

B. Results

Our method produced 2,068,894 trainable parameters, causing a training time of 15 minutes on a CPU with no GPU enabled. It was trained on 3909 samples and validated on 2143 samples. Producing an accuracy of 91% on validation data and a test score of 18%. We use 10 epochs on the training stage, the necessary to get a good generalization.

C. Conclusion

With the rebirth of neural networks and the end of the well-known winter of AI the classical methods of computer vision have become outdated. The well results of CNNs on computer vision context have proven that the use of CNNs on computer vision task is a trend with empower the computer vision to reach new levels. In this extended abstract we present a CNN architecture to classify head and shoulder from top view images as a first part of a bigger project aimed to model the human behavior in a crowd.

II. Acknowledgment

I would like to thank to SECITI in Mexico for providing the founding for the present research and the TecNM to allowed me to participate in this postdoctoral stay, also I must express my gratitude for all the team that conforms the Barcelona Super Computing Center for facilitate the equipment and the infrastructure for the project.

References

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–511–I–518, 2001. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=990517

[2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, 2005. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1467360

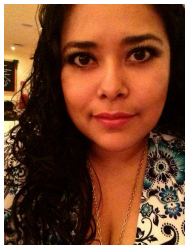
[3] P. Sermanet *et al.*, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," 2013. [Online]. Available: http://arxiv.org/abs/1312.6229

[4] A. Lee, "Comparing Deep Neural Networks and Traditional Vision Algorithms in Mobile Robotics," 2015. [Online]. Available: https://pdfs.semanticscholar.org/1b6f/569b79721037425fca034c7ae47904fb9276.pdf

[5] M. Li *et al.*, "Rapid and robust human detection and tracking based on omega-shape features," in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, nov 2009, pp. 2545–2548. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5414008

Ivan Rivalcoba received his Ph.D from Instituto Tecnolgico de Estudios Superiores de Monterrey Campus Estado de Mxico in 2015, He is a full-time professor at Instituto Tecnològico de Gustavo A. Madero in Mèxico City. Currently is working at the Barcelona Supercomputing Centre as a Postdoctoral fellow. Their work focus on using artificial intelligent methods for people detection.

Isaac Rudomin is a senior researcher at the Barcelona Supercomputer Center, which he joined in 2012. His focus is on crowd rendering and simulation including generating, simulating, animating, and rendering large and varied crowds using GPUs in consumer-level machines and in HPC heterogeneous clusters with GPUs. Previously, Isaac was on the faculty at Tecnologico de Monterrey Campus Estado de Mexico (from 1990 to 2012). He finished his Ph.D. at the University of Pennsylvania under Norman Badler on the topic of cloth modeling.



Krely Rodrguez in 2013 she completed her PhD in Mathematics Education at the Escuela Libre de Ciencias in the city of Jalapa, Veracruz. Master of Science in Electronic Engineering from the Instituto Tecnològico de Orizaba. She is currently a professor of Basic Sciences at the Instituto Tecnolgico de Minatitln, she actually collaborates in interinstitutional research projects of Artificial Vision since 2013 with Dr. Jorge Ivn Rivalcoba Rivas.